Two-layer prediction model for efficient cache uses in Cloud Computing

Jaehyun Kim, Eui-Young Chung
Department of Electrical and Electronic Engineering, Yonsei University, Korea
E-mail: jahyun96@yonsei.ac.kr

With the recent advent of the 4th industrial revolution, cloud and big data have begun to gain popularity, and the role of Solid-State Drive (SSD) is becoming important in building cloud computing servers. SSD storage has strengths in durability, reliability and high speed compared to storage configured with conventional Hard Disk Drive (HDD), but SSD is more expensive than HDD storage. To complement this, a storage structure that uses SSD as cache of HDD has been adopted to get quick response of random I/O access of cloud servers and for cost-effective manner [1]. However, due to the characteristic of cloud data, the reuse-distance is considerably long, so conventional cache algorithms such as Least Recently Used (LRU) cannot make full use of cache capabilities [2]. There was a study to distinguish reuse-distance of request using perceptron machine learning [3]. However, there could be a request with short reuse-distance (hot request), but high write ratio loaded to SSD, and we define this request as a 'dirty-hot request'. This is a problem because write request is fatal to the life expectancy of SSD. Therefore, in this paper, we propose a two-layer prediction model (Fig .1) considering both reuse-distance and write ratio to choose threshold ($\theta = write\ ratio * reuse\ distance$). If the request is over threshold($\theta_{req} > \theta_{th}$), send this request directly to HDD and if not ($\theta_{req} < \theta_{th}$), sending this request to SSD (Fig .2). First layer is used to predict reuse-distance and this result goes in to the second layer with read, write ratio. We used Openstack [4] to implement cloud computing environment and with this proposed model. With this model, it's able to determine the threshold which could leads best performance of SSD cache.

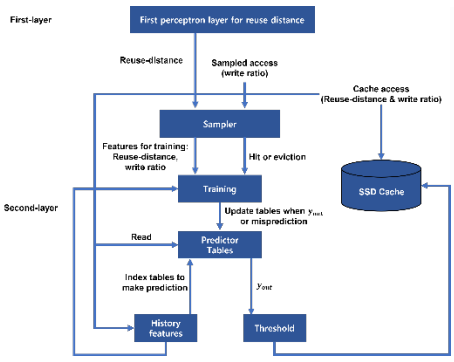

Fig 1. Two-layer prediction model for caching algorithm.

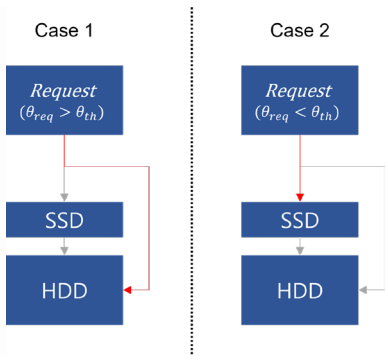

Fig 2. Cases to use storage depending on θ.

Acknowledgments This work was supported by the Samsung Electronics Company, Ltd., Hwaseong, Korea.

References

[1] Lee, D., Min, C., & Eom, Y. I. (2015). Effective flash-based SSD caching for high performance home cloud server. *IEEE Transactions on Consumer Electronics*, *61*(2), 215-221.

[2] Zhou, Ke, et al. "Efficient SSD cache for cloud block storage via leveraging block reuse distances." *IEEE Transactions on Parallel and Distributed Systems* 31.11 (2020): 2496-2509.

[3] Teran, Elvira, Zhe Wang, and Daniel A. Jiménez. "Perceptron learning for reuse prediction." *2016 49th Annual IEEE/ACM International Symposium on Microarchitecture (MICRO)*. IEEE, 2016.

[4] Openstack, https://openstack.org